DOCUMENT RESUME

ED 430 018 TM 029 731

TITLE	Contracting for Statewide Student Achievement Tests: A Review. Department of Public Instruction 98-4.
INSTITUTION	Wisconsin State Legislative Audit Bureau, Madison.
PUB DATE	1998-03-00
NOTE	41p.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Achievement Tests; *Contracts; Elementary Secondary Education; State Legislation; *State Programs; *Testing Problems; Testing Programs
IDENTIFIERS	*TerraNova Multiple Assessments; Test Publishers; *Wisconsin

ABSTRACT

The Wisconsin legislature has required the Department of Public Instruction to adopt or approve standardized tests for statewide use to measure student attainment of knowledge and concepts in grades 4, 8, and 10. Although school districts generally gave high ratings to the contents of TerraNova (McGraw Hill), the testing instrument most recently used, there was dissatisfaction with the administration of the testing program in 1996-97. A review of testing program administration, scoring, and reporting was conducted, and the test procurement process was also evaluated. Three problems were identified as leading to significant inconvenience and frustration: (1) a delay in receiving testing materials for the Milwaukee public schools with an associated delay in testing and scoring, delaying statewide results; (2) missing questions from the fourth-grade examination that required software modifications; and (3) pages that separated from test booklets, which caused scoring problems. Evaluation indicated that the testing process improved in the 1997-98 school year. Some issues in the test bidding process are considered, including problems of testing program continuity if the test vendor changes due to cost or other administrative requirements. Some changes are recommended to improve the testing program, but it is noted that frequent structural changes can diminish the program's ability to identify long-term trends. A response from the Department of Public Instruction to the State Legislative Audit Bureau is included. (Contains three tables.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 430 018

TM029731

TM

A REVIEW

Contracting for Statewide Student Achievement Tests

Department of Public Instruction

98-4

LEGISLATIVE AUDIT BUREAU

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Janice Mueller

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

A REVIEW

Contracting for Statewide Student Achievement Tests

Department of Public Instruction

98-4

March 1998

1997-98 Joint Legislative Audit Committee Members

Senate Members:	Assembly Members:
Robert W. Wirch, Co-chairperson	Mary A. Lazich, Co-chairperson
Joseph Wineke	Carol Kelso
Brian Burke	John Gard
Peggy Rosenzweig	Gregory Huber
Dale Schultz	Doris Hanson

LEGISLATIVE AUDIT BUREAU

The Bureau is a nonpartisan legislative service agency responsible for conducting financial and program evaluation audits of state agencies. The Bureau's purpose is to provide assurance to the Legislature that financial transactions and management decisions are made effectively, efficiently, and in compliance with state law and that state agencies carry out the policies of the Legislature and the Governor. Audit Bureau reports typically contain reviews of financial transactions, analyses of agency performance or public policy issues, conclusions regarding the causes of problems found, and recommendations for improvement.

Reports are submitted to the Joint Legislative Audit Committee and made available to other committees of the Legislature and to the public. The Audit Committee may arrange public hearings on the issues identified in a report and may introduce legislation in response to the audit recommendations. However, the findings, conclusions, and recommendations in the report are those of the Legislative Audit Bureau. For more information, contact the Bureau at 131 W. Wilson Street, Suite 402, Madison, WI 53703, (608) 266-2818.

State Auditor - Janice Mueller

Editor of Publications - Jeanne Thieme

Audit Prepared by

Jennifer Noyes, Director - Contact Person
Ingo Bensch

TABLE OF CONTENTS

LETTER OF TRANSMITTAL 1

SUMMARY 3

INTRODUCTION 11

Program History 12
Funding History 15

TEST ADMINISTRATION AND SCORE REPORTING 19

Testing Materials 19
Score Reports 21
Timeliness and Usefulness 21
Comparability to Prior Years 25
Contractual and Program Modifications 26
Contract Concessions 26
Program Improvements 28

PROGRAM CONTINUITY 31

Procurement of Testing Services 31
Procurement History 31
Contractual Issues 32
Additional Challenges 34
High School Graduation Test 34
Proficiency Categories 35
Timing of the Tests 36

APPENDIX - RESPONSE FROM THE DEPARTMENT OF PUBLIC INSTRUCTION

****

State of Wisconsin \ LEGISLATIVE AUDIT BUREAU

SUITE 402
131 WEST WILSON STREET
MADISON, WISCONSIN 53703
(608) 266-2818
FAX (608) 267-0410

March 23, 1998

Senator Robert W. Wirch and
Representative Mary A. Lazich, Co-chairpersons
Joint Legislative Audit Committee
State Capitol
Madison, Wisconsin 53702

Dear Senator Wirch and Representative Lazich:

We have completed a review of the Department of Public Instruction's administration of the Wisconsin Student Assessment System's Knowledge and Concepts Examinations, as directed by the Joint Legislative Audit Committee. The tests measure the achievement of fourth-, eighth-, and tenth-grade students enrolled in Wisconsin's 426 school districts in five subject areas: reading, mathematics, science, social studies, and language arts, including writing.

Although the underlying cause of problems encountered with the testing process in the 1996-97 school year was a change in testing contractors, the Department's management of the program did little to mitigate the effects of this change. In particular, the Department's lack of specificity about its expectations of the new contractor contributed to delays in the availability of score reports and concerns about their content. Tardiness by the contractor also contributed to the problems encountered. However, problems with comparability of test results to prior years could not have been avoided.

Although the contractor did not meet expectations in 1996-97, the Department renewed its contract for $1.5 million, which is an increase of $250,000 over the 1996-97 contract when the costs associated with a one-time study are excluded. The Department, however, has taken steps to improve the process during the current school year. As a result, the testing process proceeded smoothly, and score reports have been available with minor delays.

The Department has requested and received permission from the Governor to negotiate successive one-year renewals of the current contract through 2001-02, in order to ensure program continuity. Such continuity may be difficult to achieve, however, given continually changing program and policy goals. For example, tenth graders will no longer be subject to the Knowledge and Concepts Examinations process after the implementation of a high school graduation test beginning in 2000-01.

We appreciate the courtesy and cooperation of the Department of Public Instruction in completing this review. The Department's response is the appendix.

Respectfully submitted,

Janice Mueller

Janice Mueller
State Auditor

JM/JN/lw

SUMMARY

Section 118.30, Wis. Stats., requires the Department of Public Instruction to adopt or approve standardized tests for statewide use to measure pupil attainment of knowledge and concepts in the fourth, eighth, and tenth grades. The tests, which are commonly known as Knowledge and Concepts Examinations and are part of the Wisconsin Student Assessment System, measure student achievement in reading, mathematics, science, social studies, and language arts, including writing. School districts are required to administer the tests to all enrolled fourth-, eighth-, and tenth-grade students, except those whose parents or guardians have requested exemptions. Accommodations, which can include exclusion from the tests, may be made for students enrolled in a special education program and for limited-English speaking students.

The Department's Office of Educational Accountability oversees the Knowledge and Concepts Examinations program. However, rather than provide testing services itself, the Department contracts with a vendor to develop, print, distribute, and score the tests and to develop and disseminate individual and summary score reports for individual students, schools, and school districts, as well as statewide results. Since 1992-93, the first school year in which the tests were administered, the Department has contracted with three different vendors to provide testing services. The third of these vendors—CTB/McGraw-Hill—first provided services in the 1996-97 school year at a total cost to the Department of $1.425 million in general purpose revenue.

Although school districts generally gave high ratings to the contents of *TerraNova*, the testing instrument used, there was dissatisfaction with administration of the Knowledge and Concepts Examinations program in 1996-97. At the request of the Joint Legislative Audit Committee, we completed a review of administration of the program, including the provision and quality of testing materials and the content and comparability of score reports. We also reviewed the process through which *TerraNova* was procured, including whether it was appropriate. In completing our evaluation, we examined the Department's files related to the contracts for testing services; interviewed staff in the Department, representatives of CTB/McGraw-Hill, assessment coordinators in several school districts, and state procurement officials; and reviewed relevant state statutes and administrative code regarding the procurement of goods and services.

When the Department entered into the contract with CTB/McGraw-Hill, the company had not yet completed development of assorted testing

materials and score reports the Department was purchasing. The vendor took longer than anticipated to complete the development of these materials, which limited the amount of time available to customize them in order to meet Wisconsin's needs. In addition, the Department was not initially clear about some of its expectations, such as demographic information to be collected for the first time in 1996-97 to meet federal requirements under Title I of the Improving America's Schools Act, and its expectation that separate pre-test guides be developed for each of the participating grades. Together, these factors contributed to problems encountered during the testing period, including errors in and delayed distribution of the test guides, booklets, and directions.

From a review of the problems encountered with administration of the tests in 1996-97, it is apparent that they led to a significant amount of inconvenience and frustration for school districts and classroom teachers. Nevertheless, many problems were ultimately of little consequence to the testing program. However, three problems contributed to delays in the test scoring and reporting process:

- because of the delays in receiving testing materials, Milwaukee Public Schools delayed its administration of the test, resulting in the late return of approximately 16,600 tests for scoring, which delayed compilation of statewide test results;

- two questions were missing from the fourth-grade test, which required CTB/McGraw-Hill to modify its software in order to exclude these questions from the scoring process; and

- pages became separated from test booklets, which required CTB/McGraw-Hill to match pages manually with individual student booklets before scoring the tests.

The delays in the reporting of scores were of greater consequence to the school districts than the problems encountered with the administration of the tests because it is through the score reports that the Department and school districts assess the quality of education in each school and school district, identify and correct curricular weaknesses, and provide guidance to individual students. For 1996-97, score reports consisted of scale scores, which reflect the number of correct answers of each student, and national percentile rankings in each subject for each student, as well as summaries of student performance for each school, each district, and the state.

Although the Department and school districts expected the score reports to be available in January and early February, as they had been in prior years, score reports for 1996-97 were distributed over a six-month period, from February 1997 through July 1997. In addition to problems encountered in the administration of the tests, two other factors contributed substantially to the delays in the scoring process. First, as was the case with its development of supplementary testing materials, CTB/McGraw-Hill did not design its standardized reports on schedule. For example, the vendor did not provide copies of the report formats until October 23, 1996, even though the vendor's schedule indicated the Department would receive copies for review by July 15, 1996. Discussions about the customization of score reports did not begin until November 1996, two months before reporting was initially expected to begin.

Second, although the Department had envisioned a variety of adjustments to CTB/McGraw-Hill's standardized reports, it had not been sufficiently clear about its expectations in its contract with the vendor. In particular, the Department's request for proposals (RFP) for testing services, which was included as an attachment to the contract, included only a list of anticipated score reports labeled "example," although sample reports from the prior contractor were made available. The vendor's response, which was also included as an attachment to the contract, referred only to "score reports specified by the [Department]." No further clarification of expectations was included in the contract language. As a result, there were significant differences in the understanding of what changes needed to be made to the vendor's standard reports, which led to a negotiation process about the extent to which the reports would be customized to meet Wisconsin's needs that lasted through March 1997.

In addition to contributing substantially to delays in the availability of score reports, the Department's lack of clarity about its expectations also led to the production of reports that did not fulfill its or school districts' expectations. In particular, the reporting packages did not summarize schools' results by test question in order to allow a school district to assess its strengths and weaknesses on each item being tested, or identify the criteria by which student answers were graded to allow schools to adjust their writing instruction and ensure those components that are evaluated are covered adequately in class. The Department had expected CTB/McGraw-Hill to provide this information because it had been considered part of the assessment process, the prior vendor had produced these reports, and item analyses had been included in the sample list of reports in the RFP. After negotiating with the Department, CTB/McGraw-Hill subsequently provided similar, but less specific, reports for school district use.

Despite the difficulties encountered, the Department took no action to adjust the amount paid to CTB/McGraw-Hill in 1996-97 for the testing services it provided, nor did it invoke the penalties authorized under the terms of the contract. While the Department did request the removal and replacement of the CTB/McGraw-Hill manager with responsibilities for the project, staff in the Department state that contract reductions were not warranted because the vendor provided services beyond those specified in the contract. They also state that invocation of penalties would not have been prudent, given the Department's interest in maintaining an ongoing relationship with the vendor, which is reflected in the fact that the Department agreed to renew the contract with CTB/McGraw-Hill for 1997-98. Excluding $175,000 for a one-time study, the contract increased by 20 percent from $1.25 million to $1.5 million. Department staff believe the contractual increase was necessary to compensate CTB/McGraw-Hill adequately and retain the firm as the Department's testing contractor.

The testing process has improved in the 1997-98 school year. Testing materials were available on time and free of significant defects, and score reports have been available earlier than in 1996-97. Most school districts received test results for individual students and scores as scheduled in January 1998, but the Department announced a ten-day delay in the availability of score reports for districts, which was originally scheduled to begin February 13, 1998. According to the Department, shipping of these reports was completed February 27, 1998. In addition, statewide results were released in March, with the full statewide report expected to be available in April. These improvements result, in part, from the progress made throughout 1996-97 in further development of CTB/McGraw-Hill's *TerraNova* testing materials, scoring process, and customization of materials to meet the Department's expectations. In addition, the Department and CTB/McGraw-Hill have taken steps to improve contract management in 1997-98, including agreeing to a single project schedule.

Although some operational difficulties may be inevitable during the first year of a complex contract with a new vendor, the extent of the problems encountered during the 1996-97 school year raises questions about the Department's oversight and management of its contract for testing services. Therefore, we have recommended the Department improve its contract management by establishing performance-based contracts that include clear project goals and deadlines. Payment to the contractor should be contingent upon the timely fulfillment of these contractually established goals in order to ensure the services provided meet the Department's expectations.

While the Department's project management contributed to problems with both test administration and the distribution and content of score

reports, problems related to comparability with test results from prior years could not have been avoided, because the complete comparability of test results requires continuity in the tests being used. That is, the selection of a new vendor—and thereby a new test—predetermined that test results would not be comparable with those of previous years. Because problems with continuity of the test results and other problems that were encountered in 1996-97 could have been avoided if the Department had not entered into a new contract for the provision of testing services, questions have been raised about why the change in vendors was made.

The Department must follow the same general procedures as most other state agencies when purchasing goods and services. Under these guidelines, which are determined by the State Bureau of Procurement and detailed in the State Procurement Manual, most purchases of $10,000 or more—$25,000 or more for some agencies—require formal procedures for selecting a vendor, including the rebidding of a contract for professional services every three years. The Department's contract for testing services falls under these guidelines and, therefore, needed to be rebid for services beginning in the 1996-97 school year. In reviewing the vendor-selection process, we found the Department complied with state purchasing requirements. The contract was awarded following the RFP process, which allowed the Department the flexibility to specify criteria in addition to price when selecting the vendor. The Department's selection criteria for 1996-97 did not, however, explicitly include test continuity.

If test continuity is not considered in the selection criteria specified in the RFP, the need to rebid the contract every three years may undermine the Department's ability to ensure continuity within the Knowledge and Concepts Examinations program. To address concerns about the frequent change in vendors, the Department has requested and received approval from the Governor for an exemption from the next scheduled procurement process in 1998-99. This exemption will permit successive one-year extensions of the Department's contract with CTB/McGraw-Hill through 2001-02. It should be noted, however, that while it will enable the Department to secure the benefits of program continuity, it creates other risks. For example, it may limit the Department's ability to ensure it is getting the best services possible at the most competitive price.

In addition, even if the Department continues its contract with CTB/McGraw-Hill for several years, the testing program's continuity will not be assured. Three ongoing initiatives may result in significant changes to the program. First, under 1997 Wisconsin Act 27, school districts must administer a high school graduation test beginning in 2000-01 and require passage of the test, or of an alternative assessment, for all high school graduates beginning in 2002-03. Tenth-grade students

will no longer be required to be tested as part of the Knowledge and Concepts Examinations program after 2000-01. However, the high school graduation test, which will be designed to qualify individual students for high school graduation, may not also be able to provide assessment data for schools, districts, and the state that are consistent with data provided by the Knowledge and Concepts Examinations.

Second, score reports for 1997-98 include, for the first time, evaluations of students' performance in comparison to proficiency standards adopted by the Department. These standards define what fourth-, eighth-, and tenth-grade students ought to know and translate these content standards into minimum expected proficiency scores in each of four levels: advanced, proficient, basic, and minimal performance. This method of reporting differs substantially from the method used in the past, which compared students in Wisconsin only with one another and with students nationwide. Two different concerns about the Department's new proficiency categories have been expressed:

- The Department adopted its proficiency standards for the Knowledge and Concepts Examinations in November 1997, before the Governor's Advisory Taskforce on Education and Learning completed its statewide model academic standards. This has led some to question whether the Department's standards are aligned appropriately with the statewide standards. That question is significant because under 1997 Wisconsin Act 27, each district must adopt academic standards by August 1, 1998. If the standards adopted by school districts differ substantially from the Department's proficiency standards, then school districts will teach students based on one set of standards while testing them on another.

  This would become even more problematic if the current proposal to end "social promotion"—which is included in 1997 Senate Bill 436 and its companion, 1997 Assembly Bill 768—is enacted. Under this proposal, before students may advance to grades five and nine they would be required to score at the "basic" level of performance on the Knowledge and Concepts Examinations for the fourth and eighth grades or to complete satisfactorily an alternate examination adopted by a school board. In order to enable students to perform at the basic level, school district academic standards must be aligned with those the tests are designed to measure.

- Some school district administrators and others are concerned that the change to proficiency standards will reflect poorly on Wisconsin's schools, because comparisons to proficiency standards can provide a less-favorable assessment than comparisons to test results of other students across the country. Initial results indicate that in some districts, few students are performing at a level the Department deems proficient.

  In response to these concerns, 1997 Assembly Bill 642 would authorize a school board to exempt itself from the Knowledge and Concepts Examinations and, therefore, avoid receiving test results. The bill would also allow a school board to adopt a resolution exempting itself from the requirements regarding the administration of a high school graduation test. However, enactment of the bill could jeopardize Wisconsin's receipt of aid payments under Title I of the Improving America's Schools Act, which requires states to make use of high-quality assessments to determine adequate yearly progress in participating schools in at least three levels of proficiency and to identify how well children are learning. These payments are expected to total $128 million in fiscal year 1997-98.

Third, the Department has decided to move the Knowledge and Concepts Examinations to the spring semester beginning in the 1998-99 school year. The Department changed the test schedule in response to criticism that tests administered at the beginning of the year cannot measure learning that takes place during the year. However, at the school, district, and statewide levels, the change in the test schedule could complicate comparisons of some scores for the first two years of *TerraNova* testing with scores for subsequent years. It may also limit the usefulness of the scores for individual students, because test results may not be provided until late April or early May.

Although some changes to the pupil assessment program can improve assessment and accountability of Wisconsin's educational system, frequent structural changes can diminish the program's ability to identify long-term trends and contribute to administrative challenges. The Legislature and the Department will need to weigh the benefits of any changes to the Knowledge and Concepts Examinations program that might be considered in the future against the advantages of a stable and consistent program.

****

INTRODUCTION

Statewide, standardized tests measure student knowledge in the fourth, eighth, and tenth grades.

Section 118.30, Wis. Stats., requires the Department of Public Instruction to adopt or approve standardized tests for statewide use to measure pupil attainment of knowledge and concepts in the fourth, eighth, and tenth grades. The tests, which are commonly known as Knowledge and Concepts Examinations and are part of the Wisconsin Student Assessment System, measure student achievement in reading, mathematics, science, social studies, and language arts, including writing. School districts are required to administer the tests to all enrolled fourth-, eighth-, and tenth-grade students, except those whose parents or guardians have requested exemptions. Accommodations, which can include exclusion from the tests, may be made for students enrolled in a special education program and for limited-English speaking students.

School districts use test results to monitor their success in teaching students, to identify strengths and weaknesses to be addressed through curriculum adjustments, and to provide guidance and feedback to individual students and their parents. The Department also relies on the tests to fulfill a statutory requirement that it identify and assist low-performing schools, and to fulfill federal requirements under Title I of the Improving America's Schools Act that the states use high-quality assessments both to determine adequate yearly progress in participating schools in at least three levels of proficiency and to identify how well children are learning. The use of test results to determine general or categorical aids to school districts or to evaluate, hire, fire, or discipline teachers is, however, prohibited under s. 118.30, Wis. Stats.

The Department contracts with national testing firms for testing services.

Since statutory requirements regarding the tests were established in 1991 Wisconsin Act 269, the Department has contracted with three national testing firms to supply the testing instruments; to print, distribute, and score the tests; and to provide related materials. The most recent of these contracts, which is with CTB/McGraw-Hill, went into effect on July 2, 1996, for the 1996-97 school year. In October 1996, each of Wisconsin's 426 school districts administered CTB/McGraw-Hill's testing instrument, which had been newly completed and was marketed as *TerraNova*, to its fourth-, eighth-, and tenth-grade students.

School districts generally gave high ratings to the content of the *TerraNova* tests; however, there was dissatisfaction with the administration of the Knowledge and Concepts Examinations program during the 1996-97 school year. The availability of testing materials and score reports was delayed, some tests and related materials were defective, score reports lacked analytical information that had been

available in past years, and test results could not easily be compared to the results of other tests given in prior years. To address these concerns, the Joint Legislative Audit Committee directed the Legislative Audit Bureau to review:

- the administration of the Wisconsin Student Assessment System's Knowledge and Concepts Examinations during the 1996-97 school year, including the provision and quality of testing materials and the availability of test results; and

- the process through which the Knowledge and Concepts Examinations were procured, including whether it has been appropriate.

In completing our review, we examined the Department's files related to contracts for these tests. We interviewed staff in the Department, representatives of CTB/McGraw-Hill, assessment coordinators in several school districts, and state procurement officials. We also reviewed relevant state statutes and administrative code regarding the procurement of goods and services.

Program History

The Legislature established the current testing program in 1992.

Before the Legislature established the Wisconsin Student Assessment System's Knowledge and Concepts Examinations in 1992, several different assessment programs provided information about student achievement, and the purposes of the programs varied. For example:

- From 1974 until 1987, the Wisconsin Pupil Assessment Program tested randomly selected samples of elementary, middle school, and high school students in several grades. Students were tested first in reading and later also in writing, mathematics, science, economics, and government. The results of these tests were used to assess the adequacy and efficiency of educational programs in Wisconsin's public schools.

- Between 1985 and 1992, the Competency-Based Testing Program measured the proficiency of elementary, middle school, and high school students in school districts that chose to participate in the program. Proficiency was measured against minimum standards for reading, language arts, and mathematics in those districts. School districts used test results to assess individual students' progress and to identify deficiencies in educational programs. Test results also were provided to parents, but not to the Department or the public.

- Beginning in 1985, state law has required school districts to administer annually a standardized reading test developed by the Department to all third-grade students enrolled in school districts. Results from the test, known as the *Wisconsin Reading Comprehension Test*, are used to determine whether a student is a candidate for remedial reading services.

- From 1988 until 1992, state law required districts that chose not to participate in the Competency-Based Testing Program to test students in reading, language arts, and mathematics at least twice during the time period from kindergarten through grade five, at least once during the time period from grades six through eight, and at least once during the time period between grades nine through eleven, using curriculum-based achievement tests.

The first Knowledge and Concepts Examinations were administered to eighth- and tenth-graders on a voluntary basis during the 1992-93 school year; 379 of 427 school districts participated. In the subsequent school year, the eighth- and tenth-grade tests became mandatory. A voluntary fourth-grade test was added during the 1995-96 school year; 380 of 417 school districts participated. The fourth grade test became mandatory during the 1996-97 school year. As reflected in Table 1, since the testing program began, the Department has contracted with three different vendors to develop, print, distribute, and score the Knowledge and Concepts Examinations, as well as to develop score reports about the performance of individual students, schools, school districts, and the state. In each case, the Department followed the appropriate procurement process in contracting for testing services.

Table 1

Wisconsin Student Assessment System
Program History

School Year	Grades Tested: Fourth	Eighth	Tenth	Vendor	Test
1992-93		V	V	American College Testing	*Plan, Explore*
1993-94		M	M	Harcourt Brace	SAT-8
1994-95		M	M	Harcourt Brace	SAT-8
1995-96	V	M	M	Harcourt Brace	SAT-8
1996-97	M	M	M	CTB/McGraw-Hill	*TerraNova*
1997-98	M	M	M	CTB/McGraw-Hill	*TerraNova*
1998-99*	M	M	M	CTB/McGraw-Hill	*TerraNova*

V = Voluntary
M = Mandatory
* anticipated

The testing program has undergone several changes.

In addition to changes in the vendor, several other changes have been made to the Knowledge and Concepts Examinations program since 1992. These changes were made to reflect new funding requirements under Title I of the Improving America's Schools Act, which provides aid to assist disadvantaged students, as well as to strengthen Wisconsin's accountability standards and accommodate changes in the testing instruments being used. For example:

- emphasis has increased on compliance with the guidelines for the testing of students with exceptional educational needs and limited English proficiency, and the manner in which these students' scores are incorporated in school and district summary scores has changed, in order to ensure more complete participation in state assessment programs, to facilitate evaluation of programs to serve these students, and to increase the accountability of school districts for the education of all students;

- proficiency scores, which categorize students at various achievement levels based on statewide expectations for each subject included in the tests, are being reported for the first time in the 1997-98 school year; and

- proficiency standards will replace the use of national percentile rankings as criteria in the Department's identification of low-performing schools in the 1997-98 school year.

In 1995 Wisconsin Act 27, the Legislature repealed and eliminated funding for another scheduled change in the State's assessment program that would have evaluated students based on more subjective performance assessments, which could have included, for example, teacher evaluations of performance in class and student portfolios. Some believe this kind of assessment provides more valid results than standardized achievement tests, but performance-based assessment is also more expensive, time consuming, and less efficient for statewide or district-wide comparisons. Between fiscal year (FY) 1992-93 and FY 1994-95, the Department received state funds totaling more than $1.5 million for development of this program, which was to be implemented in the 1996-97 school year.

Funding History

Contract costs per student tested have increased 17.2 percent since 1992.

As shown in Table 2, expenditures for the Knowledge and Concepts Examinations increased from $591,400 in FY 1992-93, when 91,621 students in two grades were tested on a voluntary basis, to $1,425,000 in FY 1996-97, when 188,371 students participated in mandatory testing for the fourth, eighth, and tenth grades. Therefore, the cost per student increased $1.11, or 17.2 percent, from $6.45 to $7.56, since testing began. These expenditures reflect only the cost of the vendor contract for testing services, which includes development, printing, distribution, and scoring of the tests; the development and dissemination of individual and summarized score reports by individual student, school, and school district, as well as statewide; and the development and administration of a career interest survey for eighth- and tenth-grade students.

Table 2

Knowledge and Concepts Examinations
Budget and Expenditures

	General Purpose Revenue	
School Year	Budget	Expenditures
1992-93	$ 625,000	$ 591,400
1993-94	851,600	816,700
1994-95	1,110,300	934,875
1995-96	950,000	948,000*
1996-97	1,540,000	1,425,000
1997-98	1,680,000	1,500,000**
1998-99	1,690,000	n/a

* The Department supplemented this amount with $235,000 in federal Title I funds to administer the voluntary fourth-grade tests.

** Contracted amount

The Department incurs additional costs to administer the tests, including staff salaries and fringe benefits, travel, printing, and meeting costs. For example, in FY 1996-97, the Department's Office of Educational Accountability, which oversees the Knowledge and Concepts Examinations program, supported 13 full-time equivalent (FTE) positions, of which 2.5 FTE positions were assigned to the program, and expended $1,064,000 in general purpose revenue and $19,000 in federal funds on program operations. In addition to its work on the Knowledge and Concepts Examinations, the Office's expenditures also supported other programs and initiatives, such as:

- test development, standard-setting, administration, production, and reporting of the *Wisconsin Reading Comprehension Test*;

- Wisconsin's participation in the National Assessment of Educational Progress, which is a congressionally mandated survey of educational achievement administered by the U.S. Department of Education's National Center for Education Statistics; and

- coordination of a variety of assessment and accountability initiatives with school districts and within the Department, including implementation of federal Title I requirements, state assistance to low-performing schools, and assessment literacy training.

In FY 1997-98, the Office also began development of the high school graduation test required by 1997 Wisconsin Act 27, for which the Legislature appropriated $500,000 in FY 1997-98 and $850,000 in FY 1998-99.

****

Several problems were encountered during the 1996-97 testing process.

During the 1996-97 testing process, the Department and school districts encountered numerous problems with the quality and timeliness of the testing materials, the contents and timeliness of the test results, and the comparability of these results to data from prior years. Given the periodic change in contractors and the fact that operational difficulties may be encountered during the first year of a complex contract with a new vendor, it is reasonable to expect the Department to take steps to ensure the transition to a new contractor proceeds as smoothly as possible. However, the Department's management of the testing program did little to mitigate the effects of the most recent change in contractors, which were exacerbated by the contractor's delayed development of the testing materials. Of particular concern is the Department's lack of specificity about its expectations of the new contractor regarding score reports. Improvements are, however, evident in the current school year, which is the second year of the Department's contract with CTB/McGraw-Hill.

Testing Materials

CTB/McGraw-Hill had not fully developed the testing materials to be used in Wisconsin when the Department entered into a contract with the company to purchase them. Subsequently, the vendor took longer than anticipated to complete development of its standard materials and, after they were developed, to customize the materials as requested by the Department. In addition, the Department was not initially clear about some of its expectations, such as demographic information to be collected for the first time in the 1996-97 school year to meet Title I requirements and separate pre-test guides to be provided for each of the participating grades. Together, these factors contributed to delays in the distribution of testing materials to school districts and to the number of defects in the materials during the 1996-97 school year.

Testing materials were distributed late and contained some errors.

The delays in the distribution of materials that were experienced, as well as some of the errors contained in the materials themselves, included the following:

- The vendor did not provide final versions of test directions for teachers or testing guides until late August and September 1996, although the Department expected the materials by July. Consequently, the Department needed to curtail its review of drafts, final copies, and printer's proofs of the materials to ensure quality and approve content. It was later determined that some of the materials, including the test directions for fourth- and tenth-grade teachers, contained errors.

- School districts did not receive test booklets and instructions until one week before the statewide testing window of October 7–25, 1996, even though the Department's delivery schedule had called for distribution of these materials nearly two weeks before the testing window was to begin. Because the testing materials were not presorted by school, as expected, there were further delays in distribution of the materials by large districts to their individual schools.

- The tests that were distributed to the school districts contained defects. In particular, all fourth-grade test booklets were missing a page containing two test questions, and test booklet covers containing demographic information tore unexpectedly at their perforations.

Although the problems reported in 1996-97 led to a significant amount of inconvenience and frustration for school districts and classroom teachers, many of the problems were ultimately of little consequence to the testing program. However, according to department and vendor staff, three of the problems were of significant consequence because they contributed to delays in the test scoring and reporting process. In particular:

- Because of the delay in receiving the testing materials, Milwaukee Public Schools extended the testing window one week beyond the statewide period and returned the testing materials to the contractor for scoring approximately two weeks after the deadline. One high school and one elementary school postponed testing by several additional weeks. This, in turn, delayed scoring of tests for 16,600 students tested in that district and the compilation of results for the entire state.

- The two questions missing from the fourth-grade tests were ultimately dropped, which required CTB/McGraw-Hill to modify its software in order to score the tests. These adjustments also contributed to a delay in the scoring and reporting process.

- Pages became separated from test booklets, which slowed the scoring process further, because CTB/McGraw-Hill staff needed to match manually some pages containing student demographic data with the correct test booklets.

The delays in the reporting of scores were of greater consequence to the school districts than the problems encountered with the administration of the tests because it is through the score reports that the Department and school districts assess the quality of education in each school and school district, identify and correct curricular weaknesses, and provide guidance to individual students.

Score Reports

Score reports were provided late, and some did not meet expectations.

Score reports consist of scale scores, which reflect the number of correct answers for each student, and national percentile rankings in each subject for each student, as well as summaries of student performance for each school, each district, and the state. During the 1996-97 school year, three concerns were expressed about the score reports provided by CTB/McGraw-Hill. Two concerns—the score reports were not provided in a timely fashion, and they did not include all of the information to which school districts had become accustomed—were attributable to CTB/McGraw-Hill's delays in development of its standardized "off-the-shelf" score reports and the Department's initial lack of specificity concerning the expected contents of reports and related materials. The third concern—the score reports were not comparable to those of prior years—was attributable to the change in the testing instrument that occurred in 1996-97.

Timeliness and Usefulness

Two of the primary concerns about score reporting in 1996-97 were that the reports were provided significantly later than in previous years and that they excluded information that had been available with the test results provided by the previous contractor. In past years, these reports had been available in January and early February, thereby allowing school districts sufficient time to use them for mid-year curriculum adjustments and guidance counseling, including discussions about an

individual student's course selections in the subsequent school year. However, school districts were unable to make full use of score reports in 1996-97 because the score reports were received several weeks—in some cases, months—later than in prior years and excluded two reports that had been previously available.

The contractor's proposed report formats were three months late.

The development of score reports was delayed initially by CTB/McGraw-Hill's tardiness in providing standardized reports that could then be customized for Wisconsin. In particular, CTB/McGraw-Hill did not provide copies of these report formats to the Department until October 23, 1996, even though the company's schedule indicated that the Department would receive copies for review by July 15, 1996. Therefore, discussions between the Department and CTB/McGraw-Hill concerning customization of score reports did not begin until November 1996; a final list of reports to be sent to school districts was not developed until December 1996, one month before reporting was initially expected to begin; and negotiations concerning the degree to which the reports could be modified and the customization that resulted from these discussions extended through March 1997.

Differences in expectations about the score reports surfaced after CTB/McGraw-Hill submitted copies of its standardized reports to the Department for its review. The Department, although it had agreed to use the vendor's standard reports during the 1996-97 school year, had envisioned a variety of adjustments to the reports, including disaggregation of data by demographic categories under Title I; revisions to the data fields to be included in the reports; and changes in formatting, column headers, and accompanying text. However:

- CTB/McGraw-Hill's reporting software allowed only a small degree of variation from its standardized reports, such as modification or suppression of some reporting titles and score fields;

- CTB/McGraw-Hill had not budgeted programming personnel, time, or funds for modification of its software to meet the Department's expectations; and

- CTB/McGraw-Hill had not initially collected all of the demographic data that would have been required for disaggregation of results by the categories specified by Title I.

The contract lacked adequate specificity in defining the content of score reports.

These differences in expectations were the result of insufficient clarity in the contract, which incorporated both the Department's request for proposals (RFP) for testing services and CTB/McGraw-Hill's contract proposal. The RFP included a list of anticipated score reports that had been labeled "example," although sample reports from the prior contractor were made available. In turn, CTB/McGraw-Hill's proposal referred only to "score reports specified by the [Department]," without identifying which reports would be produced. The lack of further specificity in the ensuing contract resulted in differing expectations concerning both the types of reports to be produced and the degree to which the "off-the-shelf" reports that CTB/McGraw-Hill uses for its customers nationwide would be customized. This resulted in a significant, and apparently unanticipated, investment of time by both the Department and the contractor to reconcile expectations.

While negotiations about the content of the reports proceeded, other difficulties were encountered that also contributed to the delays. For example, CTB/McGraw-Hill:

- abandoned its plans to scan answers to multiple-choice questions into its computer because the format of the company's test booklets proved to be incompatible with its imaging system. As a result, the vendor hired additional staff to double-enter student answers manually into its computer.

- needed to correct inventory and demographic information submitted by school districts that was either incorrect or not useful. For example, many school districts omitted information required on the inventory sheets that accompanied the test materials, and some students used different variations of their names on the test booklets and demographic information pages.

- did not complete development of national empirical norms, which are used in score reporting to compare student performance nationwide, until February 1997.

The Department did not keep school districts informed about delays.

The lack of clarity about expectations, coupled with the vendor's operational difficulties, affected the usefulness of the test results in two ways. First, reports were available significantly later than in previous years. As shown in Table 3, score reports for individual students, schools, and districts were distributed over a two-month period, from late February 1997 through April 1997, compared to distribution of all reports by early February in previous years; statewide results were not

disseminated until July 1997. School district frustration with the delay in reporting was exacerbated by poor communication from the Department, which had initially indicated score reports would be completed sometime in February. The Department did not provide updated or more specific reporting schedules until March 1997, and the actual delivery of reports occurred up to two weeks later than indicated in this updated schedule.

Table 3

Delivery Dates for Score Reports

Reports	Delivery Dates in Prior Years	1996-97 Actual Delivery Dates
Student and School Reports	January 8 – 9	February 25 – March 21
District Reports	January 8 – 9	April 4 – 15
Statewide Reports	January 23 – February 2	July 11 – 15

It appears that the Department was unable to provide accurate information to the districts about the report distribution schedule because it fully expected CTB/McGraw-Hill to produce the test results on the same time schedule the previous vendor had followed in prior years; the Department's RFP for the testing contract included a reporting schedule that required the provision of all test results in January 1997. Although both the Department and CTB/McGraw-Hill subsequently produced more detailed project schedules, differences between the two schedules were not reconciled, so that there was no single project schedule governing the testing process in 1996-97. In addition, neither schedule was formally updated to reflect contract negotiations that resulted in an increase in the number of test questions to be scored and a change in the fourth-grade testing period from spring to fall. However, even if there had been agreement about a schedule, it is unlikely that such a schedule would have been met or would have enabled the Department to provide accurate reporting schedules to the districts, given the delays in the availability of CTB/McGraw-Hill's "off-the-shelf" score reports and the substantial investment of time required to customize these reports to meet the Department's expectations.

Second, even after the score reports were issued, they did not provide all of the information school districts had expected, based on reporting packages available from the prior vendor. In particular, CTB/McGraw-Hill:

- had not scheduled the production of item analyses, which summarize a school's results by test question and are used as a tool through which the school can identify its strengths and weaknesses within each subject being tested, until the second year of its production schedule; and

- declined to issue task-specific scoring rubrics for the writing tests, which identify the criteria by which student answers are graded and are a tool through which schools can adjust their writing instruction to ensure components that are evaluated are covered adequately in class, because it considers rubrics to be proprietary.

The Department had expected the vendor to produce item analyses and scoring rubrics. The Department's expectations were based on the development of such materials by the previous vendor, the fact that it shared copies of these materials with CTB/McGraw-Hill and other vendors during the procurement process, and the inclusion of item analyses in a sample list of reports in the Department's RFP. However, representatives of CTB/McGraw-Hill believed the Department had accepted the *TerraNova* package of reports, which did not include item analyses during the first year or scoring rubrics for the writing tests. Ultimately, the Department negotiated with CTB/McGraw-Hill to produce other, less specific reports to summarize school performance by groups of questions and samples of student essays for each possible score on the writing test.

Comparability to Prior Years

The third major concern about the test results for 1996-97 was that they could not easily be compared to prior years' results, which impaired the Department's and school districts' ability to identify trends in student achievement. However, given that a new testing instrument was being used for the first time, there was little the Department could do to achieve comparability of results.

The Department took steps to address concerns about comparability.

Nevertheless, the Department did include in the contract with CTB/McGraw-Hill the requirement that a study be completed to allow for limited comparisons between the scores of the tests administered from 1993-94 through 1995-96 and the *TerraNova* test administered in 1996-97. Although the study that CTB/McGraw-Hill had initially proposed was not conducted because of anticipated administrative difficulties, an alternative study involving a statistical linking procedure

was done. Department staff believe this change, which was suggested by the Department, improved the reliability of the data produced by this linking study.

However, although a linking study was completed, the information produced by it can be used only for general comparisons at the school and district level. The linking study does not allow comparisons across test years for the scores of individual students, small groups of students, or the entire population of Wisconsin students who took the test. In addition, the Department advises school districts against using the linking study for any significant policy or programmatic decisions.

Contractual and Program Modifications

Despite the difficulties encountered by the Department in its management of the Knowledge and Concepts Examinations during the 1996-97 school year, it took no action to adjust the amount it paid to CTB/McGraw-Hill for the provision of testing services, which totaled $1.425 million. According to staff in the Department, although not all expectations were met, the vendor made other, substantial contributions to the testing program for which it did not receive remuneration. However, it is not clear whether the value of these additional services adequately compensated for the concessions made by the Department during the 1996-97 school year. CTB/McGraw-Hill's performance and the testing process itself have improved during the 1997-98 school year.

Contract Concessions

The lack of clarity about expectations in the 1996-97 contract required the Department and CTB/McGraw-Hill to define expectations throughout the contract's term. While the negotiation process allowed some of the problems encountered to be resolved efficiently and effectively, it also resulted in contract concessions. For example, the Department:

- accepted reductions in the degree to which score reports and other testing materials were customized;

- allocated its own staff to develop the 1996-97 edition of the *Wisconsin School Performance Report* and contracted with Badger State Industries for printing services, at an additional cost of $7,000, although the Department had expected CTB/McGraw-Hill to produce this report; and

- accepted the late delivery of services, tests, score reports, and other material and did not assess penalties of $1,000 per day authorized by the contract for such late delivery. Given the ambiguity of the contract language, the amount that could have been assessed ranges from $84,000 to the maximum of $712,500, depending on how the contract's provisions are interpreted.

The Department did request the removal and replacement of the CTB/McGraw-Hill manager with responsibilities for the project. However, staff in the Department state that invocation of penalties would not have been prudent, given the Department's interest in maintaining an ongoing relationship with the vendor. In addition, staff believe there was no reason to make adjustments to the contract because CTB/McGraw-Hill agreed to provide services not specified in the contract, to which the Department and vendor assigned an estimated value exceeding $150,000. These services included two one-day assessment literacy training workshops valued at more than $50,000 for district and department staff, and provision of 500 free copies of CTB/McGraw-Hill's teacher's guide to *TerraNova*, which are normally sold for $25 a copy. However, other services that staff in the Department consider to be additional could reasonably be expected to be included under the terms of the original contract, including the contractor's development of a customized data tape containing state-level summary information and a disaggregated report of student scores by various demographic categories.

In addition, although staff in the Department assert the vendor exceeded its contractual obligations in its level of participation in a six-day proficiency score standard-setting process in April 1997, at which nearly 200 educators defined proficiency categories for each grade level and subject included on the Knowledge and Concepts Examinations and assigned ranges of scores to each category, it does not appear the services provided were beyond the scope of the initial agreement. In particular, CTB/McGraw-Hill's response to the RFP described significant involvement by the company in arranging logistics, facilitating the standard-setting process, and preparing materials and reports. The only service that CTB/McGraw-Hill provided that appears to have exceeded these contractual obligations is voluntarily reimbursing participant expenses, such as lodging, travel costs, and meals, totaling approximately $55,000.

Despite the Department's assertion that it was more than compensated for the concessions it made through CTB/McGraw-Hill's provision of additional services, the total value of these services has not been compared to the value of the Department's concessions. In addition, the

value assigned to these services by CTB/McGraw-Hill may differ from their actual costs and from the amount CTB/McGraw-Hill would charge in a competitive environment. Therefore, whether the Department ultimately received services comparable to those it intended to purchase for the contracted amount of $1.425 million is subject to interpretation. It is evident, however, that the Department did not receive the services it had expected, regardless of whether other, substitute services were subsequently provided by the vendor.

Program Improvements

School districts report improvements in the testing program during the current school year.

The Department and school districts report improvements in CTB/McGraw-Hill's performance during the administration of the tests in 1997-98. Materials for the pre-test workshops and testing materials arrived on or before the scheduled delivery dates. In addition, the test booklets and instructions were packaged for easy distribution to the schools and appeared to be free of any significant errors or defects. Because of these improvements—which resulted largely from CTB/McGraw-Hill's progress in completing development of its testing materials and customization of these materials to meet the Department's expectations—the Department and districts reported generally smooth administration of the tests.

The Department and CTB/McGraw-Hill also have taken steps to improve the reporting phase of the testing process. For example:

- reporting schedules were provided to school districts at their pre-test workshops held in September 1997, and the Department has communicated revisions in these schedules to the districts;

- test booklets were modified to allow electronic imaging of student responses;

- the scoring software was revised to improve the match between test booklets and pages containing demographic information of students who used different variations of their names, and to ignore some stray marks on the demographic page;

- item analyses for all questions, which summarize a school's results by test question in order to allow it to identify strengths and weaknesses within each subject being tested, will be included as part of the reporting package to be sent to all school districts in 1997-98;

- the renewal contract between the Department and CTB/McGraw-Hill included a list of reports to be prepared by CTB/McGraw-Hill; and

- although not specifically included in the contract, the Department has specified the information to be included in the customized reports and communicated these requirements to CTB/McGraw-Hill.

The Department and CTB/McGraw-Hill also included a single, comprehensive project schedule in their renewal contract, which was intended to ensure timeliness and reduce the potential for miscommunication. The schedule identifies numerous interim steps, such as development of draft test materials and score reports, and establishes deadlines for these interim steps and final delivery of testing materials and score reports to school districts and the Department. CTB/McGraw-Hill and the Department have adhered to most parts of the schedule, but they have missed some deadlines by several weeks.

For example, although score reports are being sent to school districts earlier in 1997-98 than they were in 1996-97, there have been some delays. CTB/McGraw-Hill's shipping logs indicate that score reports for individual students and schools were distributed to 327 of the 426 districts by January 16, 1998, which was the deadline established by the Department. Subsequently, CTB/McGraw-Hill sent reports to the remaining districts. Department staff attribute delays in the reports to errors in demographic and inventory information provided by the districts, including Milwaukee Public Schools, which submitted incorrect data for as many as 1,500 students.

In addition, the Department delayed the reporting schedule for district reports by 10 days, in part because heavy rains in California caused CTB/McGraw-Hill to close its offices and delay score processing. These reports were scheduled for release beginning February 13, 1998, but, according to the Department, shipping was not completed until February 27, 1998. This delay was also expected to affect the release of statewide information, which had already been rescheduled in December 1997 because the initially anticipated release date of February 22, 1998, proved to be unrealistic. Statewide results were released in March, with the full statewide report expected to be available in April.

Despite the improvements in 1997-98, it is evident that the process did not proceed as planned during the 1996-97 school year. In addition, the Department did not receive any financial compensation or reduction in the contracted amount from the contractor. Although the Department believes that, through negotiations, the vendor provided services beyond

those required in the contract, such negotiations would not have been necessary if expectations had been adequately specified in the contract. Therefore, in order to minimize the extent to which problems with the Knowledge and Concepts Examinations program are encountered in the future and to ensure the Department has clear contractual authority to link payments to the contractor's performance, *we recommend the Department of Public Instruction's future contracts for testing services include clear interim and final project goals and deadlines, and make payment of the contracted amount contingent upon timely and successful completion of the contractually established project goals.*

****

The change in the testing vendor has been questioned.

Because the problems encountered during the 1996-97 administration of the Knowledge and Concepts Examinations could have been avoided had the Department not entered into a contract with a new vendor, questions have been raised about why the change in vendors was made. The change was precipitated by state procurement guidelines, which require that most professional services contracts be rebid every three years. Given the unique nature of the Knowledge and Concepts Examinations program, including the need to ensure comparability of test results over a period of years, the appropriateness of the procurement guidelines in this situation can be questioned. While steps have already been taken to ensure program continuity in the future, consideration also needs to be given to several other factors that may affect the future of the testing program.

Procurement of Testing Services

Most state agencies must follow the same general purchasing procedures, which are determined by the State Bureau of Procurement within the Department of Administration and detailed in the State Procurement Manual. Purchases of $10,000 or more—$25,000 or more for some agencies—require formal procedures for selecting a vendor and usually require prior authorization from the Bureau of Procurement. In reviewing the vendor-selection process, we found the Department complied with state purchasing requirements. The contract was awarded following the RFP process, which allowed the Department the flexibility to specify criteria in addition to price when selecting the vendor. While the Department's selection of CTB/McGraw-Hill was in compliance with the established procedures, the need to rebid the contract every three years undermines the Department's ability to ensure continuity within the Knowledge and Concepts Examinations program.

Procurement History

The Department did not initially follow the RFP process in securing a contract for testing services. Rather, in 1992-93, the first year of the Knowledge and Concepts Examinations program, the Department sought and received a sole-source waiver to enter into a contract with American College Testing, Inc. (ACT) to develop tests of student knowledge and to score the tests. Sole-source contract waivers are requested when the Department determines only one vendor can appropriately supply needed materials or services. If a waiver request is approved, the Department is

not required to follow the normal procurement process of obtaining price quotes from more than one source. However, a detailed written justification supporting the need for a sole-source contract must be prepared by the Department and approved by the State Bureau of Procurement. The Governor must also approve these waivers.

The Department's waiver from bidding requirements for its contract with ACT would have allowed for continuity in the testing program by permitting the Department to enter into five consecutive one-year contracts. However, concerns about the contract resulted in reviews by the departments of Administration and Employment Relations. While these reviews did not report any violations of state purchasing guidelines, after only one year the Department of Administration revoked its approval of the waiver allowing for continued renewal of the sole-source contract with ACT.

Since the Department of Administration's decision to revoke the sole-source waiver, the Department of Public Instruction has issued an RFP for testing services twice, in 1993 and in 1996. On each occasion, the Department of Public Instruction has awarded the contract to a different vendor, leading to substantial modifications in the testing program. The selection of CTB/McGraw-Hill in spring 1996 was based on the Department's evaluation of the *TerraNova* tests as superior to those of the existing contractor, which was the only other bidder, but continuity of the tests being used was not an explicit selection criterion. However, the problems encountered in 1996-97 led the Department and some school districts to call for an extension of the current contract with CTB/McGraw-Hill in order to ensure program continuity in the future.

The Department has received approval to extend the current contract through 2001-02.

The Department recently requested—and the Governor approved—an exemption from the next scheduled procurement process in 1998-99. The Governor's approval in January 1998 of a sole-source waiver will permit successive one-year extensions of the Department's contract with CTB/McGraw-Hill through 2001-02. Department staff indicate they intend to request permission from the Department of Administration to extend the contract cycles for the Knowledge and Concepts Examinations to five or more years when they issue their next RFP in 2001-02. The Department believes this should ensure continuity in the program and help to avoid the problems encountered in 1996-97.

Contractual Issues

While longer contracting cycles will enable the Department to secure the benefits of program continuity, the Department will need to consider the risks of foregoing frequent procurement processes if it chooses continual renewal of its current contract with CTB/McGraw-Hill. First,

consideration needs to be given to whether the Department's decision to extend is contractual relationship with CTB/McGraw-Hill was premature, given the problems encountered in 1996-97. The contractor has met performance expectations regarding test administration during the current school year, yet some deadlines regarding score reports have been missed. Nevertheless, if the contractor's performance is adequate overall, some of these concerns will be alleviated. However, the Department needs to consider, on an annual basis, whether continued renewal of the current contract is the most appropriate strategy to follow for the program, based on the contractor's performance.

It is instructive that the Department, after following the RFP process on two different occasions, chose to change vendors primarily because the new vendors offered tests that more closely matched the Department's specifications. For example, staff in the Department state *TerraNova* was purchased because its content was aligned with the Department's emerging model academic standards.

Negotiating a competitive contract price may be difficult.

Second, the Department needs to consider how it can best negotiate competitive contract amounts during annual renegotiations with the current contractor without testing the market to determine potential vendors' best offers. In contrast to simple contract renewals, the procurement process provides agencies with an opportunity to ensure they are paying competitive prices for the goods and services they purchase from contractors. For example, the Department's initial contract with CTB/McGraw-Hill in 1996-97 was based on a bid submitted by CTB/McGraw-Hill in competition with another firm, so the Department had some assurance that CTB/McGraw-Hill made its best offer.

The Department's negotiations with CTB/McGraw-Hill for renewal of its contract for 1997-98 did not have the benefit of such price comparisons. This may have been detrimental to the Department's bargaining position when CTB/McGraw-Hill requested an increase in the contracted amount. Although the new contract increased the cost of CTB/McGraw-Hill's services only 5.3 percent, from $1.425 million in 1996-97 to $1.5 million in 1997-98, the effective increase is $250,000, or 20 percent, because the 1996-97 contract included $175,000 for a one-time study that has been completed.

Questions can be raised about recent contract cost increases.

Given the absence of any significant new contract requirements, the rationale for the increase in the contract is not apparent. Representatives of the Department and CTB/McGraw-Hill cited inflation and continued customization in the format of the test materials and score reports as reasons for the increase, but some of the customization is completion of the work the Department had expected in 1996-97. Based on our review, the only significant new contractual expectation appears to be the inclusion of proficiency reports. In addition, staff in the Department

stated they believed the Department would need to pay more to retain the company as its contractor, but, the Department has no documented information about CTB/McGraw-Hill's actual costs or the extent to which the company's costs will decrease in subsequent contract years after start-up costs decrease.

In considering the appropriate length of contract cycles, the departments of Public Instruction and Administration will need to weigh the benefits of greater program continuity against the benefits of frequent procurement processes. One alternative to a reduction in the frequency of procurement processes might be the issuance of RFPs that take into consideration the benefits of continuity by including test consistency as a scoring criterion, along with other content factors and cost. In addition, the Department needs to scrutinize annual requests for contractual increases more carefully to ensure they can be justified by inflation, changes in contract requirements, or changes in market forces. A comparison of contract expenditures by other states for similar testing services provided by the same or competing vendors might offer one indication of fair market price.

Additional Challenges

Changing policy and program goals threaten program continuity.

To facilitate effective use of test results from the Knowledge and Concepts Examinations, the Department will need to maintain stability in the program. However, program continuity may be difficult to achieve as policy and programmatic goals continue to drive changes to the program and affect the comparability of test results. Concerns have been raised about three ongoing initiatives that may result in significant changes to the program in the next several years.

High School Graduation Test

Tenth-grade students will not be subject to the Knowledge and Concepts Examinations after 2000-01.

1997 Wisconsin Act 27 established a high school graduation test in place of the current tenth grade Knowledge and Concepts Examinations. The Act requires school districts to administer a high school graduation test beginning in 2000-01 and to require passage of the test, or of an alternative assessment, for all high school graduates beginning in 2002-03. In addition, the Act eliminates the requirement that school districts administer Knowledge and Concepts Examinations to tenth-grade students after 2000-01.

However, the high school graduation test may not be able to qualify individual students for high school graduation and provide assessment data for schools, districts, and the state that is consistent with Knowledge and Concepts Examinations, because the two tests will serve different

purposes. In contrast to the Knowledge and Concepts Examinations, the graduation test:

- may differ among school districts, because the statutory provision gives school districts a choice between establishing their own graduation tests or administering a test being developed by the Department;
- will be administered at various grade levels; and
- can be taken multiple times by students who do not receive passing scores.

Proficiency Categories

Changes to score reporting went into effect during the current school year.

Score reports for 1997-98 include, for the first time, evaluations of students' performance in comparison to proficiency standards adopted by the Department. These standards define what fourth-, eighth-, and tenth-grade students ought to know and translate these content standards into minimum expected scores on the Knowledge and Concepts Examinations for students in each of four levels of proficiency: advanced, proficient, basic, and minimal performance. This method of reporting differs substantially from the method used in the past, which compared students in Wisconsin only with one another and with students nationwide. CTB/McGraw-Hill included arbitrary objective performance estimates with the 1996-97 test results, but these scores were not based on any content-based proficiency standards or other analysis by the Department or school districts. Two different concerns about the Department's new proficiency standards have been expressed.

First, the Department adopted its proficiency standards for the Knowledge and Concepts Examinations in November 1997, before the Governor's Advisory Taskforce on Education and Learning completed its statewide model academic standards. This has led some to question whether the Department's standards are aligned appropriately with the statewide standards. The question is significant because under 1997 Wisconsin Act 27, each district must adopt pupil academic standards in mathematics, science, reading and writing, geography, and history by August 1, 1998. If the standards adopted differ substantially from the Department's proficiency standards, then school districts would teach students based on one set of standards while testing them on another.

This will become even more problematic if the current proposal to end "social promotion" is enacted. The proposal, which is included in

1997 Senate Bill 436 and 1997 Assembly Bill 768, would require that as a prerequisite to advance to grades five and nine, students must score at the "basic" level of performance on fourth- and eighth-grade Knowledge and Concepts Examinations or satisfactorily complete an alternate examination adopted by a school board. In order to enable students to perform at the basic level, school district academic standards must be aligned with those the tests are designed to measure.

Second, some school district administrators and others are concerned that the change to proficiency standards will reflect poorly on Wisconsin's schools. In particular, because comparisons of test results to proficiency standards are expected to provide a less-favorable assessment than comparisons to test results nationwide, some have raised concerns about whether Wisconsin's students and schools will be unfairly penalized for the change in score reporting. Initial results indicate that, in some districts, few students are performing at a level the Department deems proficient. In response to these concerns, 1997 Assembly Bill 642, which has been referred to the Assembly Committee on Education, would authorize a school board to exempt itself from the Knowledge and Concepts Examinations and, therefore, avoid receiving test results and participating in the State's accountability program. The bill would also allow a school board to adopt a resolution exempting itself from the requirements regarding the administration of a high school graduation test.

While enactment of 1997 Assembly Bill 642 would allow for local control over the administration of student proficiency tests, it could also jeopardize the State's aid payments under Title I of the Improving America's Schools Act. Aid payments under Title I, most of which the Department distributes to school districts to offset a broad range of expenditures that assist disadvantaged students, are expected to total $128 million in FY 1997-98. However, a requirement of receiving the aid payments is that recipients conduct annual pupil assessments at three grade levels to demonstrate yearly progress of schools and local educational agencies that receive aid funds.

Timing of the Tests

The tests will be given in the spring, rather than the fall, beginning in 1998-99.

Finally, the Department has decided to move the Knowledge and Concepts Examinations to the spring semester beginning in the 1998-99 school year. As a result, tests in 1998-99 will not be administered until the end of February and the beginning of March. Previously, the Knowledge and Concepts Examinations have been administered in the fall semester, with two exceptions. The eighth- and tenth-grade tests in 1993-94 were delayed to the spring after the Department was required to complete a new procurement process, and

the voluntary fourth-grade test in 1995-96 was held in the spring after funding and a testing instrument became available for this expansion of the Knowledge and Concepts Examinations program.

The Department changed the test schedule in response to criticism that tests administered at the beginning of the year cannot measure learning that takes place during the year, even though the Department has consistently maintained that the tests measure students' cumulative learning through the test date rather than learning specific to the year the tests are administered. Previously, the Department had already removed the grade level from the name of the tests to counter the perception that the tests are intended to measure learning in any particular grade level.

At the school, district, and statewide level, the change in the test schedule could complicate comparisons of scores for the first two years of the *TerraNova* test with scores in subsequent years. Department staff indicate CTB/McGraw-Hill has established national normative comparisons based on both fall and spring administration of the tests, which will allow comparisons of national percentile rankings for tests administered in either the fall or spring. However, comparisons of scale scores, which reflect the number of correct answers for each student, or proficiency scores, which are based on scale scores, may be less meaningful because students will benefit from several additional months of instruction.

In addition, spring testing will delay the availability of score reports, which will reduce their benefit to students. As noted, schools use individual students' scores in guidance counseling, including selection of appropriate classes for the following year. Districts have indicated they need to receive scores in the beginning of the spring semester to make full use of the test results, but staff in the Department anticipate that scores will not be available until late April or early May. Unexpected delays, such as those experienced during the current school year, could further postpone reporting.

Although some changes to the pupil assessment program can improve assessment and accountability of Wisconsin's educational system, frequent structural changes can diminish the program's ability to identify long-term trends and contribute to administrative challenges. The Legislature and the Department will need to weigh the benefits of any changes to the statewide standardized assessment program that might be considered in the future against the advantages of a stable and consistent program.

****

State of Wisconsin
Department of Public Instruction
Mailing Address: P.O. Box 7841, Madison, WI 53707-7841
125 South Webster Street, Madison, WI 53702
(608) 266-3390 TDD (608) 267-2427 FAX (608) 267-1052
Internet Address: www.dpi.state.wi.us

John T. Benson
State Superintendent

Steven B. Dold
Deputy State Superintendent

March 17, 1998

Janice Mueller
State Auditor
Legislative Audit Bureau
131 West Wilson Street
Madison, Wisconsin 53703

Dear Ms. Mueller:

This letter is in response to your review of DPI's administration of the 1996-97 Wisconsin Student Assessment System's Knowledge and Concepts examinations. You correctly point out some of the difficulties CTB/McGraw-Hill and DPI encountered in the first year of contract administration. DPI's goal in this contract is to provide quality assessments for Wisconsin students, parents, teachers, and school districts that fulfill legislative requirements. Despite administrative problems in the first year of the contract, the TerraNova testing product achieves this goal. In fact, we believe TerraNova is a superior assessment instrument for Wisconsin students. We regret the audit did not discuss more extensively the quality of the assessments themselves.

The audit review also provides an historical summary of assessment activities in Wisconsin. A focus in the review is the increase in cost over a period of years. We wish to emphasize certain factors contributing to the changing cost. The first year assessments were administered, testing was voluntary at the eighth and tenth grades, and not all districts participated. In 1996, testing was added at grade 4. As more districts and students were tested, costs increased. At first, tests were entirely multiple choice, with a hand-scored writing sample. Today, the tests are more comprehensive and integrate both multiple-choice and constructed-response questions. As the tests have become more sophisticated and the scoring more complex, the costs of administration have risen.

From DPI's perspective, the administration of the assessment program requires a balance among maintaining continuity of the program over time, achieving outstanding contractor performance, and getting the best value possible for the expenditure of state dollars. The TerraNova test score reports and related services were still in development when the contract was awarded, but the test instrument has proven to be of very high quality. DPI and the contractor have made significant adjustments in the contract to meet the needs of Wisconsin school districts. As the audit explains, improvements occurred once DPI and the contractor worked through the difficulties in the early phases of the contract term. Nevertheless, we will comply with the LAB recommendation that, " . . . future contracts for testing services include clear interim and final project goals and deadlines, and make payment of the contracted amount contingent upon timely and successful completion of the contractually established project goals."

On page 8, the audit confuses the issues of content standards and proficiency scoring. It is appropriate, however, for the audit to raise the issue of the alignment of the test items to the content standards. DPI has paid close attention to the alignment of the content standards with the fourth-, eighth-, and tenth-grade tests. We have determined that nearly all the TerraNova items align with the final standards. The department intends to improve the alignment still more as future tests are developed.

DPI recognizes the need, in annual contract negotiations, to assess cost increases in relation to current market factors in order to assure the state pays competitive rates. Working with the same contractor and testing instrument over a period of several years will be beneficial to the administration of both the tests and the contract deliverables. The second-year improvements are an excellent example of how contract administration improves with experience between the parties. The approval by DOA Procurement and the Governor's office for successive contract renewals is a critical component in establishing continuity and a high-quality assessment process in Wisconsin.

In summary, I believe the department's record is strong with respect to management of responsibilities related to student assessment. But, the LAB is correct in identifying shortcomings in our administration of the 1996-97 contract. We will certainly work to improve our performance. DPI and CTB/McGraw-Hill have tried to be responsive to legislators' concerns about the 1996-97 test administration experience and its implications for future test administration, for example, by providing information and briefings to legislators and legislative committees. We appreciate legislative oversight and interest in the program and believe this will result in a more effective Wisconsin educational accountability system.

Sincerely,

John T. Benson
State Superintendent